REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	May 1996	Professional Paper

4. TITLE AND SUBTITLE

SMOOTHING DISJOINT FORMANT TRACK BOUNDARIES CAUSED BY WAVE-FORM SUBSTITUTION IN PACKET VOICE COMMUNICATION

5. FUNDING NUMBERS

PR: CE77
PE: 0603013N
WU: DN302104

6. AUTHOR(S)

E. VonColln and V. Goncharoff

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Command, Control and Ocean Surveillance Center (NCCOSC)
RDT&E Division
San Diego, CA 92152–5001

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Naval Engineering Logistics Office
P. O. Box 2304, Eads Street
Arlington, VA 22201

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This paper describes a method for smoothing disjoint formant track boundaries that can arise in packet voice communication. The disjoint boundaries can occur when missing packets are reconstructed using waveform substitution methods. Our algorithm attempts to match formant locations to LPC-poles and then adjusts the pole locations across the boundary to smooth the formants at the boundary. We describe our method for matching formants to poles and our method for adjusting the poles. The LPC residual is not modified. Time domain plots and LPC spectrograms are used to show that the formant track was smoothed. Perceptually subjective listening tests confirm that an improvement in the speech quality is achieved with this method in that it sounds more clear and natural.

Published in *Proceedings of the 6th International Conference on Signal Processing Applications and Technology*, Vol. II, pp. 1836–1840, October 1995.

14. SUBJECT TERMS

Mission Area: Communication
Voice Packet Communication
Waveform Substitution
Disjoint Formant Tracks

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAME AS REPORT

21a. NAME OF RESPONSIBLE INDIVIDUAL	21b. TELEPHONE *(include Area Code)*	21c. OFFICE SYMBOL
E. VonColln	(619) 553–3655	Code 44213

SMOOTHING DISJOINT FORMANT TRACK BOUNDARIES CAUSED BY WAVEFORM SUBSTITUTION IN PACKET VOICE COMMUNICATION

Eric VonColln

NCCOSC RDT&E Division
Code 44213
53245 Patterson Rd
San Diego, CA 92152-5000
e-mail: voncolln@cod.nosc.mil

Vladimir Goncharoff

University of Illinois at Chicago
Dept of EECS, Mail Code 154
851 S. Morgan St
Chicago, IL 60607-7053
e-mail: goncharo@eecs.uic.edu

ABSTRACT

In this paper we describe a method for smoothing disjoint formant track boundaries that can arise in packet voice communication. The disjoint boundaries can occur when missing packets are reconstructed using waveform substitution methods. Our algorithm attempts to match formant locations to LPC-poles and then adjusts the pole locations across the boundary to smooth the formants at the boundary. We describe our method for matching formants to poles and our method for adjusting the poles. The LPC residual is not modified. Time domain plots and LPC spectrograms are used to show that the formant track was smoothed. Perceptually subjective listening tests confirm that an improvement in the speech quality is achieved with this method in that it sounds more clear and natural.

1. INTRODUCTION

Voice packet communication refers to the speech coding technique where sampled speech is broken up into packets, encoded, transmitted and decoded at the receiver. Usually these packets are 8-48 msecs in duration. It is possible that the receiver misses packets due to noise bursts, transmission errors, or congested channels. Therefore the need arises to reconstruct the missing packets from the surrounding packets. Many techniques have been developed to compensate for the missing packets [1, 9, 5]. The simplest method involves no reconstruction and simply inserts silence for any missing packets. This has been shown to cause noticeable artifacts in the received speech, and is disturbing to listeners. Another solution is waveform substitution, in which the missing packet is estimated from the previous packet (and possibly the succeeding packet). Waveform substitution can involve pattern matching, pitch waveform replication, and packet merging [4, 6, 10]. Some of these methods try to smooth the time discontinuity at the boundaries between the missing packet and the two it adjoins.

The waveform substitution method is very efficient and works well as long as the speech is relatively static during the missing packet. However if significant changes occur in the speech during the missing packet the resulting speech can contain noticeable acoustical artifacts and can lead to unnatural sounding speech. We develop a method that can improve the disjointness that arises at the boundaries of missing packets when the speech is changing quickly. This method attempts to smooth a speakers formants across these boundaries by warping the LPC poles on both sides of the boundary toward a midpoint. This will insure a smooth pole transition, which, if the poles are mapped appropi-

ately to the speakers formants, will result in a smooth formant track.

2. METHOD

In this example we used a pitch waveform replication method [4] to reproduce the missing packet of speech. This leads to an abrupt formant change, as seen in Figure 3, which we now try to compensate for using our algorithm. Given two short speech segments $s1$ and $s2$ on either side of the disjoint boundary we compute LPC-9 coefficients [8, 7, 2] on windowed speech segments producing two LPC matrices $S1LPC$ (M x N1) and $S2LPC$ (M x N2). Here M is the LPC order and N1 and N2 are the number of frames in $s1$ and $s2$. Each column represents the LPC vector used for that frame of speech. Also the residual matrices used for synthesis are computed, but these are not altered. From the LPC matrices the poles are calculated as the LPC polynomial roots, forming the two matrices $S1PL$ and $S2PL$ each containing the same number of frames as its corresponding LPC matrix.

Next the poles of the last frame of $s1$ (last column of $S1PL$) are paired with the poles of the first frame of $s2$ (first column of $S2PL$) in such a way that the poles correspond to the formant locations. This is done by first sorting the poles in each frame by angle then assigning each pole as a formant if the pole magnitude is greater than 0.8. This magnitude was found empirically and has worked well for finding formants. So the first pole in the sorted vector that has magnitude greater than 0.8 is labeled as the first formant, the second pole that has magnitude greater than 0.8 is labeled as the second formant, and so on. The algorithm will try to find four formants. Any pole not labeled as a formant is not altered in the warping process. And if two frames that are matched have differing numbers of pole formants the lower number is used and the formants not mapped are not altered.

Once the poles are matched to formants a mid-point between the two pole vectors $p1$ and $p2$, where $p1$ is the last column of $S1PL$

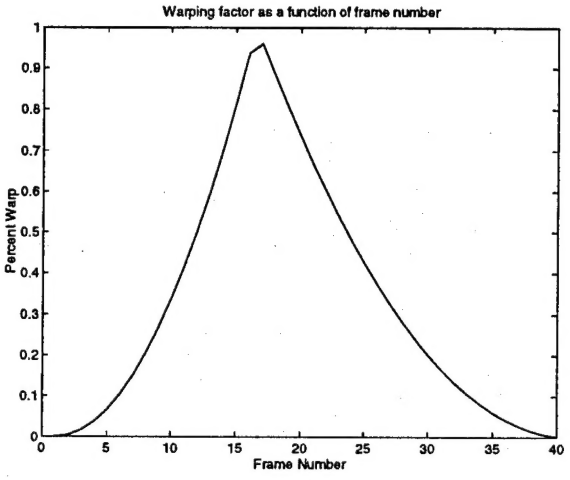

Figure 1: Percent warp of poles toward midpoint pole vector as a function of frame number.

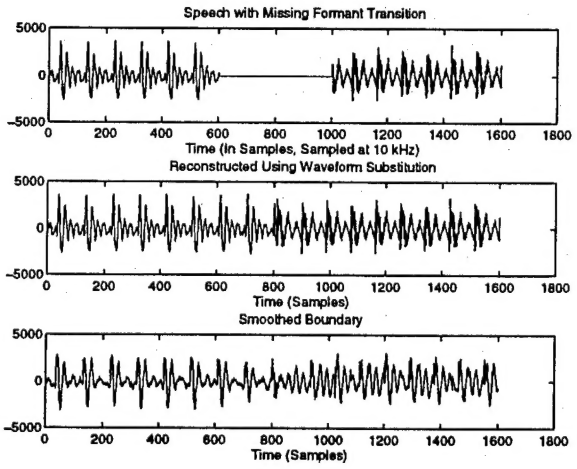

Figure 2: Speech waveforms showing waveform substitution and formant smoothing.

and $p2$ is the first column of $S2PL$, is calculated in phase and magnitude as

$$pmid_i = \left(p1_i * \frac{p2_i}{p1_i}\right)^{0.5}$$

for each pole in the vector that corresponds to a formant. Here i indicates the formant number. Then each frame of poles in $S1PL$ and $S2PL$ is recomputed by warping them towards the mid-point poles as [3]

$$pnew_{i,j} = \left(p_{i,j} * \frac{pmid_i}{p_{i,j}}\right)^a$$

where j=frame number,

$$a = \left(\frac{i-1}{N1-0.5}\right)^f$$

for $p = p1$, and

$$a = \left(\frac{N1+N2-i}{N2-0.5}\right)^f$$

for $p = p2$. Here $pnew$ is the new pole matrix and f is a simple scaler value that determines the shape of the warping factor a. The amount of warp is a nonlinear function that depends on the distance the frame is from the boundary. The warping function for $f = 2$ can be seen in Figure 1. This value has been determined empirically through subjective listening tests. The closer the frame is to the boundary the more it is warped to the mid-point poles so that at the boundary the poles from the two segments will have been completely warped to the mid-point pole vector. In Figure 1, N1=16, N2=24, and the mid-point occurs at the boundary, frame number 16, where the warp is almost 1. So, e.g., frame 10 will be warped approximately 22% towards the mid-point pole vector.

Once all the poles have been recomputed the speech is resynthesized by computing new LPC coefficients from the new pole matrix, and resynthesizing using the old residual. The final frame of the first segment and the first frame of the second segment will now be spectrally smooth in terms of a smooth pole transition. This will correspond to a smooth formant transition if the poles are matched correctly.

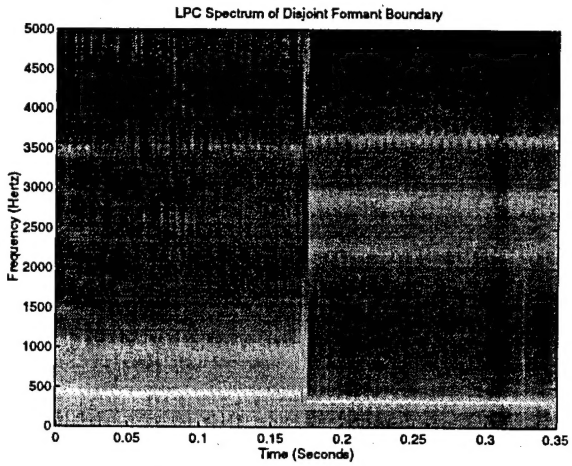

Figure 3: LPC spectrogram of disjoint formants caused by waveform substitution.

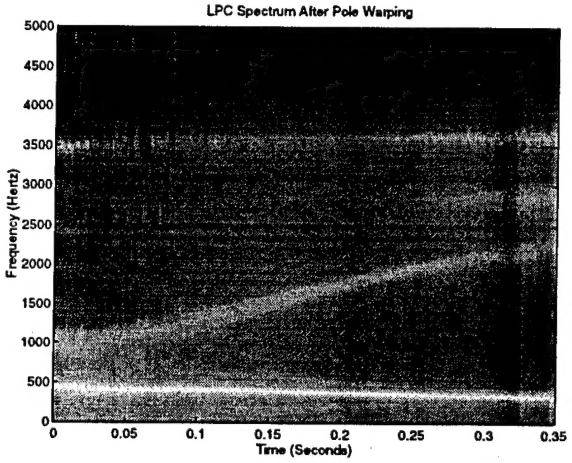

Figure 4: LPC spectrogram of boundary after pole warping.

3. RESULTS

Figure 2 shows three speech segments. The first subplot shows a speech waveform that was put through a voice packet communication system where a packet was lost. The next subplot shows the missing packet replaced with pieces from its surrounding packets in order to fill the hole. Around sample 800 we can see a very abrupt change in the waveform. It is clear that the statistics of the waveform changed very rapidly at this point. This is due to the fact that a formant change occurred in the missing packet.

We took the waveform in the second subplot and ran it through our algorithm to smooth the formants. The results can be seen in the third subplot. Note that the transition region is now much smoother. Perceptually the waveform in the third subplot sounds more natural and clear than the wavform in the second subplot.

The smoothing effects of our algorithm can be seen clearer in Figures 3 and 4. Figure 3 shows an expanded view of an LPC spectrogram of the speech waveform of Figure 2 second subplot. We use LPC spectrograms in order to highlight the formant structure. In this figure we can clearly see where the abrupt formant change occurs. This formant change has an unnatural perceptual sound quality to it. Figure 4 shows an LPC spectrogram of the speech waveform form Figure 2 third subplot. Here we can see that the formant transition across the boundary is now much smoother, especially in the second formant which had the most severe change.

4. CONCLUSION

In this paper we have developed a method of smoothing disjoint formant tracks that can occur in waveform substituted speech. It should be noted that this method does not degrade the performance of the waveform substitution methods when the speech is relatively static. This is because the formants will be relatively static and will not be warped a significant amount, if any. A good preprocessing step would be to first detect a disjoint formant, then bypass the formant smoothing if the formants are not disjoint.

The only disadvantage to this method is the increased processing time for the analysis and synthesis. Since only the boundaries, which are very short in duration, need to be smoothed our method still involves much less computation than a full synthesis technique.

5. REFERENCES

[1] J. B. Evans and T. G. Champion. Robust parametric speech coding and reconstruction techniques. In *Proceedings of ICSPAT*, pages 928–931, 1992.

[2] S. Furui. *Digital Speech Processing, Synthesis, and Recognition.* Marcel Dekker, Inc., New York, New York, 1993.

[3] V. Goncharoff and M. Kaine-Krolak. Interpolation of LPC spectra via pole shifting. In *Proceedings of ICASSP*, pages 780–783, May 1995.

[4] D. J. Goodman, G. B. Lockhart, O. J. Wasem, and W. C. Wong. Waveform substitution techniques for recovering missing speech segments in packet voice communication. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-34(6):1440–1448, Dec 1986.

[5] O. G. Jaffe. Reconstruction of missing packets of PCM and ADPCM encoded speech. Master's thesis, M. I. T., June 1986.

[6] N. S. Jayant and S. W. Christensen. Effects of packet losses in waveform coded speech and improvements due to an odd even sample-interpolation procedure. *IEEE Trans. Comm.*, COM-29(2):101–109, Feb 1981.

[7] T. W. Parsons. *Voice and Speech Processing.* McGraw-Hill, Inc., New York, New York, 1987.

[8] L. R. Rabiner and R. W. Schafer. *Digital Processing of Speech Signals.* Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1978.

[9] R. A. Valenzuela and C. N. Animalu. A new voice-packet reconstruction technique. In *Proceedings of ICASSP*, pages 1334–1336, 1989.

[10] O. J. Wasem, D. J. Goodman, C. A. Dvorak, and H. G. Page. The effect of waveform substitution on the quality of PCM packet communications. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-36(3):342–348, March 1988.